



HAL
open science

Genome Wide Association Study Pinpoints Key Agronomic QTLs in African Rice *Oryza glaberrima*

Philippe Cubry, H el ene Pidon, Kim Nhung Ta, Christine Tranchant-Dubreuil, Anne-C eline Thuillet, Maria Holzinger, H el ene Adam, Honor e Kam, Harold Chrestin, Alain Ghesqui ere, et al.

► To cite this version:

Philippe Cubry, H el ene Pidon, Kim Nhung Ta, Christine Tranchant-Dubreuil, Anne-C eline Thuillet, et al.. Genome Wide Association Study Pinpoints Key Agronomic QTLs in African Rice *Oryza glaberrima*. *Rice*, 2020, 13 (1), 10.1186/s12284-020-00424-1 . hal-03289676

HAL Id: hal-03289676

<https://hal.science/hal-03289676>

Submitted on 5 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.

1 **Title**

2 **Genome wide association study pinpoints key agronomic QTLs in African rice *Oryza glaberrima***

3 **Authors**

4 Philippe Cubry^{*1}, H el ene Pidon^{*1,2}, Kim Nhung Ta^{*3,4}, Christine Tranchant-Dubreuil¹, Anne-C eline Thuillet¹, Maria
5 Holzinger¹, H el ene Adam¹, Honor e Kam⁵, Harold Chrestin¹, Alain Ghesqu iere¹, Olivier Fran ois⁶, Fran ois Sabot¹,
6 Yves Vigouroux^{□,1}, Laurence Albar^{□,1}, Stefan Jouannic^{□,1,3}

7 **Author's affiliation**

8 ¹ DIADE, Univ Montpellier, IRD, Montpellier, France

9 ² present address: Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Seeland, Germany

10 ³ LMI RICE, AGI, IRD, Univ Montpellier, CIRAD, USTH, Hanoi, Vietnam

11 ⁴ present address: National Institute of Genetics, Mishima, Shizuoka, Japan

12 ⁵ INERA, Bobo Dioulasso, Burkina Faso

13 ⁶ Universit e Grenoble-Alpes, Centre National de la Recherche Scientifique, Grenoble, France

14 * Equally contributed to this work

15 □ Equally contributed to this work

16 **Corresponding authors:**

17 PC: philippe.cubry@ird.fr ; YV: yves.vigouroux@ird.fr ; LA: laurence.albar@ird.fr ; SJ: stephane.jouannic@ird.fr

18 **Abstract**

19 Background

20 African rice, *Oryza glaberrima*, is an invaluable resource for rice cultivation and for the improvement of biotic and
21 abiotic resistance properties. Since its domestication in the inner Niger delta ca. 2500 years BP, African rice has
22 colonized a variety of ecologically and climatically diverse regions. However, little is known about the genetic basis of
23 quantitative traits and adaptive variation of agricultural interest for this species.

24 Results

25 Using a reference set of 163 fully re-sequenced accessions, we report the results of a Genome Wide Association Study
26 carried out for African rice. We investigated a diverse panel of traits, including flowering date, panicle architecture and
27 resistance to *Rice yellow mottle virus*. For this, we devised a pipeline using complementary statistical association
28 methods. First, using flowering time as a target trait, we demonstrated that we could successfully retrieve known genes
29 from the rice flowering pathway, and identified new genomic regions that would deserve more study. Then we applied
30 our pipeline to panicle- and resistance-related traits, highlighting some interesting QTLs and candidate genes (including
31 *Rymv1* for resistance and *SPI*, *Ghd7.1*, *APO1* and *OsMADS1* for panicle architecture). Lastly, using a high-resolution
32 climate database, we performed an association analysis based on climatic variables, searching for genomic regions that
33 might be involved in adaptation to climatic variations.

34 Conclusion

35 Our results collectively provide insights into the extent to which adaptive variation is governed by sequence diversity
36 within the *O. glaberrima* genome, paving the way for in-depth studies of the genetic basis of traits of interest that might
37 be useful to the rice breeding community.

38

39 **Keywords**

40 African rice, Genome Wide Association Study, flowering time, panicle architecture, RYMV, climate variation

41

42 **BACKGROUND**

43 African rice, *Oryza glaberrima* Steud., was domesticated independently of Asian rice *Oryza sativa* L. (Wang et al.
44 2014; Meyer et al. 2016; Cubry et al. 2018; Choi et al. 2019). Its domestication took place in the inner delta of the Niger
45 river (Cubry et al. 2018), from a wild relative species, *Oryza barthii* A. Chev.. Its origin from this wild Sahelian species
46 certainly explains its strong tolerance or resistance to biotic and abiotic stresses (Sarla and Swamy 2005). In the context
47 of increasing temperatures and a more variable climate, strong tolerance to such stresses is an important objective for
48 rice agriculture worldwide. However, knowledge of the genetic basis of phenotypic variation in African rice remains
49 very limited. With the exception of salinity tolerance (Meyer et al. 2016), few association studies have been performed
50 for traits of agricultural interest in this species. Genome wide association studies (GWAS) have successfully identified
51 genes of functional importance associated with flowering time in Asian rice (Zhao et al. 2011; Huang et al. 2012; Yano
52 et al. 2016). For Asian rice, the genetic determination of this trait is well understood (Lee and An 2015), whereas we
53 have no information about the variation of this trait for African rice. Another trait of broad interest for rice farmers and
54 breeder communities is the architecture of the panicle. This trait is one of the main components of yield potential,
55 because the number of seeds per panicle is directly related to the branching complexity of the inflorescence (Xing and
56 Zhang 2010). With increasing global movement of plant material and climate change, biotic threats to rice agriculture
57 continue to evolve and the search for new sources of resistance to pathogens is therefore a challenging research field.
58 *Rice yellow mottle virus* (RYMV) is responsible for one of the most damaging diseases of rice in Africa (Kouassi et al.
59 2005; Issaka et al. 2012; Kam et al. 2013). Resistance genes against RYMV are mostly found in *O. glaberrima*, and this
60 species may be an interesting source of quantitative trait loci (QTLs) for global rice breeding strategies (Thiémmélé et al.
61 2010).

62 To better assess the functional variation present in African rice, we developed a genome-wide association panel and
63 corresponding phenotypic datasets for flowering time, inflorescence architecture, and resistance to RYMV. Using
64 several complementary statistical models for genetic association, we identified key QTLs for flowering time variation,
65 panicle architecture, quantitative resistance to RYMV and climatic variation.

66 **RESULTS**

67 A total of 892,539 SNPs from 163 different *O. glaberrima* accessions was used in the GWAS. We considered that
68 population structure was best described by four genetic groups, and we used either one or both of the groupings and the
69 kinship matrix as cofactors to correct for confounding effects in our analyses. The phenotypic data were obtained from
70 infield experiments or from available public databases (data in Additional file 1: Table S1). Using ANOVA as the
71 benchmark, all methods allowed an efficient correction for false positives (QQ-plots, see Additional files 2, 3, 4, 5).

72 **Flowering time**

73 A genome-wide association study of flowering time based on data from the early planting dates allowed us to identify
74 1,450 SNPs statistically associated with this trait (5% FDR threshold, Table 1; Additional file 2; Additional file 6: Table
75 S2). Most of these SNPs were at a distance of less than 25kb from each other and were clumped into 80 genomic
76 regions containing a total of 733 annotated genes (Table 1; Additional file 7: Table S3; Additional file 8: Table S4).
77 Corresponding analyses performed using data from the later planting date revealed associations with a slightly larger
78 number of significant SNPs: 2,649 significant SNPs corresponding to 134 regions and 1,223 annotated genes (Table 1;
79 Additional file 2; Additional file 6: Table S2; Additional file 7: Table S3; Additional file 8: Table S4). Two hundred
80 and sixty-two genes representing 25 genomic regions were found for both planting dates. Several GWAS peaks co-
81 localized with known Asian rice flowering time genes either for early or late planting (Table 2; Fig. 1): *OsGI* and
82 *OsMADS51* on chromosome 1; *Hd6/OsMADS14/OsPHYC* and *ETR2* on chromosome 3; *Hd3a/RFT1* on chromosome 6;
83 and *RCN1* on chromosome 11. When we compared the list of genes found in our study for flowering time with an
84 expert list based on bibliography (Additional file 9: Table S5), we found a five-fold significant enrichment (G test
85 associated p -value = $7e10^{-4}$), showing that our GWAS approaches were effective in retrieving known genes.

86 GO term enrichment analysis performed on the complete list of genes found five significantly (p -value < 0.05) enriched
87 GO terms using classic Fisher and weighted tests, two for cellular components (Golgi apparatus and peroxisome) and
88 three for molecular functions (motor activity, carbohydrate binding and transcription factor activity) (Additional file 10:
89 Table S6). An additional GO term was found for biological processes (Cellular processes) with the weighted test only.

90 **Panicle morphological traits**

91 Using six morphological traits, all association methods led to comparable numbers of significant associations (Table 1).
92 The CATE model produced the highest number of significant SNPs for four of the six morphological traits. A total of
93 1,010 significant SNPs was detected for at least one panicle morphological trait, except for spikelet number for which
94 no significant SNP was found (Table 1; Additional file 3; Additional file 6: Table S2). Ninety-seven unique genomic
95 regions associated with one or more morphological traits were defined (Table 1; Additional file 7: Table S3). Nine of
96 the 97 unique regions were associated with more than one trait: one for secondary branch number (SBN) and primary
97 branch average length (PBL) (rOg-PAN-18), three for primary branch number (PBN) and rachis length (RL) (rOg-
98 PAN-25, rOg-PAN-58, rOg-PAN-92), two for PBN and secondary branch internode average length (SBintL) (rOg-
99 PAN-8, rOg-PAN-56), two for RL and SbinL (rOg-PAN-78, rOg-PAN-88) and one for PBN, RL and SBintL (rOg-
100 PAN-90). The 88 remaining regions were each associated with a single trait: 32 for RL, 20 for PBN, 10 for PBL, 1 for
101 primary branch internode average length (PBintL) and 25 for SBintL. Among the 97 associated regions, 858 annotated

102 genes were identified (Table 1; Additional file 8: Table S4). GO term analysis revealed no significant enrichment of
103 gene categories in the regions of significant SNPs (Additional file 10: Table S6).

104 Some genes already known to be involved in panicle development control co-localized with genomic regions associated
105 with panicle morphological traits, examples including: *ABERRANT PANICLE ORGANIZATION1 (APO1)* which
106 colocalizes with the region rOg-PAN-59 identify for the SBintL trait; *SHORT PANICLE1 (SPI)* with the rOg-PAN-91
107 region (PBL); *OsMADS1/LEAFY HULL STERILE1 (LHS1)* with the rOg-PAN-22 region (RL); and
108 *Ghd7.1/Hd2/OsPRR37* with the rOg-PAN-65 (RL) region (Table 2).

109 **RYMV resistance**

110 Our association study of quantitative RYMV resistance detected a total of 2,199 associated SNPs (Table 1; Additional
111 file 4; Additional file 6: Table S2). These SNPs defined 257 genomic regions containing 2,248 annotated genes (Table
112 1; Additional file 7: Table S3; Additional file 8: Table S4). The EMMA model revealed a much higher number of
113 significant SNPs compared to the other models, 1,831 SNPs and 177 regions being detected only with this model.
114 Regions that were the most consistent across the different statistical methods were observed on chromosome 3 (rOg-
115 RYMV-125, position 10.5Mb, Additional file 7: Table S3), chromosome 4 (rOg-RYMV-159/rOg-RYMV-160, position
116 25-25.3Mb), chromosome 6 (rOg-RYMV-200, position 15.1Mb) and chromosome 11 (rOg-RYMV-63, position 9,0-
117 9.2Mb; rOg-RYMV-77/rOg-RYMV-78, position 26.2-26.8Mb).

118 We detected significant associations in the vicinity of two known resistance genes: *RYMV1*, located in rOg-RYMV-159
119 on chromosome 4; and *RYMV3*, located in rOg-RYMV-77 on chromosome 11 (Albar et al. 2006; Pidon et al. 2017)
120 (Fig. 2; Table 2). Analysis of GO terms (Additional file 10: Table S6) revealed enrichment in genes from the Golgi
121 apparatus, known to be involved in both the replication of some plant viruses and in intracellular trafficking and cell-to-
122 cell movement (Pitzalis and Heinlein 2018).

123 **Environment-related variables**

124 In order to study association with environmental variables, we downloaded climatic variables for 107 geolocalised
125 accessions (Cubry et al. 2018) from the worldclim v1.4 database (Hijmans et al. 2005). The first two axes of the PCA of
126 bioclimatic data (BioPC1 and BioPC2) explained 49.91% and 26.32% of the variance of all variables. BioPC1 was
127 mainly explained by bio4 (temperature seasonality) and bio17 (precipitation of driest quarter) while BioPC2 was mainly
128 explained by bio9 (mean temperature of driest quarter) and bio11 (mean temperature of coldest quarter). We evaluated
129 the statistical association of genetic polymorphisms with the first and second axes of the PCA. No significant
130 association was detected with BioPC1. BioPC2 allowed the identification of 52 SNPs (Table 1; Additional file 5;

131 Additional file 6: Table S2) and 17 regions encompassing 128 annotated genes (Additional file 7: Table S3; Additional
132 file 8: Table S4). Among the identified genes, we found a cation/proton exchanger-encoding gene (*LOC_Os02g04630*,
133 *OsCAX4*), an auxin-responsive gene (*LOC_Os03g53150*, *OsAUX*) and a MADS-box gene (*LOC_Os03g11614*,
134 *OsMADS1*) (Table 2). GO term enrichment tests detected only one GO term as significantly over-represented,
135 GO:0005576 (Cellular Components / extracellular region) (Additional file 10: Table S5).

136 When considering maximum temperature variables (Tmax), 73.56% and 17.90% respectively of the variance was
137 explained by the first PCA axis (TmaxPC1) and second PCA axis (TmaxPC2). No association was found with
138 TmaxPC1, while TmaxPC2 allowed us to detect 228 regions including 1,946 annotated genes (Table 1; Additional file
139 7: Table S3 and Additional file 8: Table S4). GO term analysis revealed several GO terms as being significantly
140 enriched, including seven terms for Biological Processes (cell differentiation, cellular protein modification process,
141 cellular component organization, embryo development, signal transduction, regulation of gene expression and response
142 to external stimulus), two terms for Cellular Components (mitochondrion and plasma membrane) and five terms for
143 Molecular Functions (receptor activity, kinase activity, nucleotide binding, translation factor activity/RNA binding and
144 transferase activity) (Additional file 10: Table S5). One genomic region, rOg-Tmax-36, co-localized with a known QTL
145 for seedling cold tolerance in *O. sativa* (Kim et al. 2014).

146 **DISCUSSION**

147 **Overlap of flowering time and panicle architecture genetic networks between African and Asian crop species**

148 The flowering pathway is a well described pathway in Asian rice *O. sativa*, with several known key genes (Tsuji et al.
149 2011; Hori et al. 2016). Based on our expert list of known flowering genes, we were able to show that our method
150 significantly retrieved genes involved in the variability of this trait. We detected candidate regions that co-localize with
151 some previously described flowering genes. Among them, *Hd3a* and *RFT1* are florigen genes homologous to
152 *Arabidopsis thaliana FT* (Komiya et al. 2008). *Hd6* is a flowering repressor, causing late flowering under Long Day
153 (LD) conditions (Ogiso et al. 2010). Overexpression of *OsMADS14*, a homolog of *A. thaliana API*, has been shown to
154 be responsible for extreme early flowering in Asian rice (Jeon et al. 2000). *OsMADS51* is a type I MADS-box gene that
155 has been shown to promote flowering in Short Day (SD) conditions (Kim et al. 2007). *OsGI* is an ortholog of the *A.*
156 *thaliana GIGANTEA* gene and its over-expression in rice leads to late flowering under both SD and LD conditions
157 (Hayama et al. 2003). *ETR2*, encoding an ethylene receptor, has been shown to delay flowering by regulating the
158 expression of *OsGI* (Wuriyangan et al. 2009; Tsuji et al. 2011). Functional characterization of the *RCN1* gene,
159 homologous to *A. thaliana TFL1*, showed that this gene can promote late-flowering when overexpressed in transgenic

160 plants (Nakagawa et al. 2002). Finally the phytochrome *OsPhyC* gene encodes one of the three rice phytochromes and
161 has been shown to be a flowering repressor (Takano et al. 2005).

162 A large number of additional genomic regions (Additional file 7: Table S3) were identified for which further studies
163 should be conducted in order to identify novel genetic diversity relating to the flowering pathway of African rice. The
164 large number of regions observed is to be expected when dealing with polygenic traits and it should be noted that a
165 comparable number of low effect QTLs was previously reported for flowering time in Asian rice on the basis of genetic
166 mapping studies (Hori et al. 2016).

167 Spikelet number per panicle was the main trait contributing to the diversity of panicle architecture observed in this
168 population. No significant SNPs were identified for the SpN trait using any of the 4 models. A possible explanation for
169 this result is that the trait might be associated with a large number of QTLs of low effect sizes, and may consequently be
170 difficult to assess using the present GWAS panel. For panicle traits, the detected peaks were broadly characterized by
171 relatively high *p*-values, suggesting that the associated morphological traits are highly polygenic with small effect sizes.
172 Some genes previously implicated in the regulation of panicle development and/or architecture in *O. sativa* (Wang and
173 Li 2011; Teo et al. 2014) were also found to be associated with panicle morphological trait variations in *O. glaberrima*,
174 indicating a parallel evolution of the trait in the two species. The *APO1* gene encodes an ortholog of the *A. thaliana*
175 UNUSUAL FLOWER ORGAN (UFO) F-box protein and was reported to be involved in the control of rice spikelet
176 number and panicle branching complexity through effects on meristem fate and cell proliferation (Ikeda et al. 2007;
177 Ikeda-Kawakatsu et al. 2009). This gene was also found to be associated with QTLs relating to yield, panicle
178 architecture complexity and lodging resistance in *O. sativa* (Terao et al. 2010; Ookawa et al. 2010). The
179 *OsMADS1/LEAFY HULL STERILE1 (LHS1)* gene encodes a *SEPALLATA*-like MADS-box transcription factor (TF),
180 which promotes the formation of spikelet/floret meristems (Jeon et al. 2000; Agrawal et al. 2005; Khanday et al. 2013).
181 Interestingly, this gene was also identified within a major drought-tolerant QTL (qDTY3.2) (Saikumar et al. 2014) from
182 a cross between *O. sativa* and *O. glaberrima*, corresponding to a genomic region associated with response to
183 environmental variables (rOg-BioClim-3). *SHORT PANICLE1 (SPI)* encodes a putative PTR family transporter, which
184 controls panicle size and branching complexity (Li et al. 2009). The *Ghd7.1/Hd2/OsPRR37* gene encodes a *PSEUDO-*
185 *RESPONSE REGULATOR (PRR)* protein associated with a QTL displaying pleiotropic effects on spikelet number per
186 panicle, plant height and heading date (Liu et al. 2013).

187 Several association studies of panicle morphological trait diversity have been recently conducted for *O. sativa* (Bai et al.
188 2016; Crowell et al. 2016; Rebolledo et al. 2016; Ta et al. 2018; Yano et al. 2019). Only a few overlaps of GWAS
189 candidates were observed between the two rice crop species. As for the flowering time analysis, a large number of
190 specific genomic regions (Additional file 7: Table S3) were identified. For these regions, further studies should lead to

191 the precise identification of genetic elements governing panicle diversity in African rice. Among the genes located in
192 these specific regions, some were previously functionally characterized in *O. sativa* but without any observations of a
193 direct relationship with panicle development, examples including: *OsIPT2* (*ADENOSINE PHOSPHATE*
194 *ISOPENTENYLTRANSFERASE*), encoding a cytokinin biosynthetic gene regulated by *OsARID3* which is essential for
195 shoot apical meristem (SAM) development in rice (Xu et al. 2015); *VERNALIZATION INSENSITIVE 3-LIKE 1*
196 (*OsVILI*) which is involved in flowering time through *OsLF* and *GhD7* genes (Jeong et al. 2016); and the *ent*-kaurene
197 synthase genes *OsKS1* and *OsKS2*, which are involved in GA biosynthesis and associated with a GWAS mega locus
198 related to panicle and yield traits in *O. sativa* (Sakamoto and Matsuoka 2004; Crowell et al. 2016). Other genes may be
199 of interest on the basis of their annotations and the known functions of their orthologs, examples including: *OsGRF5*
200 and *GRF7*, two members of the *GROWTH-REGULATING FACTOR* or GRF small gene family encoding TFs
201 associated with SAM maintenance and flowering time in rice (Omidbakhshfard et al. 2015); and *AP2_EREBP77* and
202 *OsRAV2*, two members of the *RELATED TO ABI3 AND VPI* or *RAV* subfamily of AP2/ERF TFs, homologous to the
203 *TEM* genes from *A. thaliana* that affect FT induction through the photoperiod and GA pathways (Swaminathan et al.
204 2008).

205 **Quantitative resistance to RYMV in *O. glaberrima* and major resistance genes**

206 The large number of significant SNPs associated with RYMV resistance may reflect the highly quantitative nature of
207 partial resistance. However 83% of significant SNPs were detected only by the EMMA method (1,831 SNPs),
208 suggesting that a high number of false positive SNPs was detected with this method.

209 The regions identified as being associated with resistance against RYMV did not overlap with QTLs of partial
210 resistance against RYMV previously identified in *O. sativa* (Boisnard et al. 2007), suggesting that different genes and
211 pathways may lead to resistance. However, two major resistance genes against RYMV, *RYMV1* and *RYMV3*, were
212 found to be associated with quantitative resistance. *RYMV1* encodes a translation initiation factor that acts as a
213 susceptibility factor through its interaction with the genome-linked protein (VPg) of RYMV, which is required for viral
214 infection (Albar et al. 2006; Hébrard et al. 2010). Several different alleles conferring high resistance are known
215 (Thiémmélé et al. 2010). The main candidate resistance gene for *RYMV3* belongs to the family of nucleotide-binding
216 domain and leucine-rich repeat containing (NLR) genes (Pidon et al. 2017), many of which are involved in pathogen
217 recognition and effector-triggered immunity (de Ronde et al. 2014). NLR genes are known to be frequently organized
218 into clusters and several additional NLR genes, annotated in the close vicinity of the *RYMV3* candidate gene, might also
219 be good candidates for quantitative resistance. Translation initiation factors and NLRs are known to act as determinants
220 of high and monogenic resistance (de Ronde et al. 2014; Sanfaçon 2015). Moreover, a role in quantitative resistance has
221 been clearly established for NLRs (Wang et al. 1999; Hayashi et al. 2010) and suggested for translation initiation

222 factors (Nicaise et al. 2003; Marandel et al. 2009). The *RYMV1* and *RYMV3* loci, or adjacent NLR gene loci, might thus
223 harbor both alleles with quantitative effects and alleles with strong effects on RYMV resistance. Similarly, even if the
224 *RYMV2* major resistance gene to RYMV was not found to be associated with partial resistance in this study, a previous
225 analysis strongly suggested that *RYMV2* might be involved in partial resistance to RYMV in the species *O. sativa*
226 (Orjuela et al. 2013).

227 In addition to the above, genes encoding protein domains implicated previously in virus resistance, such as lectin
228 (Yamaji et al. 2012) or methrin-TRAF domains (Cosson et al. 2010), were also found to be located in or close to
229 significantly associated regions (Additional file 8: Table S4) and may constitute interesting candidates. Further studies
230 should be conducted on these candidate genes, as well as on the additional genomic regions identified (see Additional
231 file 7: Table S3) in order to describe the diversity of resistance pathways to RYMV in African rice.

232 **Relationship between environment-related variables and *O. glaberrima* diversity**

233 We did not find any significant association between the first PCA axis and either the set of bioclimatic variables or the
234 monthly average maximal temperature. This could be explained by a high correlation between the first PCA axis and
235 population genetic structure (Frichot et al. 2015). Correcting for confounding effects did not allow us to detect
236 association at SNPs displaying allelic frequencies that correlated with such a structure. For the second PCA axis, several
237 significant associations were found and interesting genes were identified, including genes related to drought or cold
238 tolerance. Among the most interesting candidates, *OsCAX4* is a salt-tolerance gene (Molla et al. 2015), *OsAUX* is an
239 auxin-responsive gene that has been found to be significantly over-expressed in cold-tolerant seedlings (Dametto et al.
240 2015) and *OsMADSI* is located within a major drought-tolerant QTL (qDTY3.2) detected from a cross between both *O.*
241 *sativa* and *O. glaberrima* (Saikumar et al. 2014).

242 Several candidate regions have been identified that will require a more in-depth study in order to gather variations of
243 interest for breeding purposes (see Additional file 7: Table S3). This is especially true for the temperature-related
244 candidates, the latter being likely targets for genetic improvement in the context of a changing climate.

245 **CONCLUSIONS**

246 We report on the results of an extensive Genome Wide Association Study carried out for several traits on African rice.
247 Analysis of the well-studied character of flowering time enabled us to retrieve a significant number of previously
248 identified genes, thus validating our approach. We also carried out the first GWAS analysis to date of climate variables
249 in relation to African rice, obtaining a list of candidate regions that should be the subject of further studies in order to
250 detect functional variation linked to local adaptation and resistance or tolerance to abiotic stresses.

251 The RYMV resistance and panicle architecture traits for which we report significant associations are key agronomic
252 traits that should interest farmers and breeders. Further studies aiming at identifying adaptive polymorphisms among the
253 candidates found in this study and functional validation of candidate genes will be needed to reinforce our results.

254

255 MATERIAL AND METHODS

256 Genotypic data

257 Single nucleotide polymorphisms (SNPs) from 163 high-depth re-sequenced *O. glaberrima* accessions were used in this
258 study (Cubry et al. 2018). SNPs were identified based on mapping to the *Oryza sativa japonica* cv. Nipponbare high
259 quality reference genome in terms of assembly and annotation (Kawahara et al. 2013). The bioinformatic mapping
260 pipeline, software and SNP filtering steps that were used are described in Cubry et al. (2018).

261 SNPs with more than 5% missing data (minor fraction of total SNP set) were filtered out (Cubry et al. 2018). As
262 missing data can reduce the power of association studies (Browning 2008; Marchini and Howie 2010), we imputed the
263 remaining missing data based on a matrix factorization approach using the “impute” function from the R package LEA
264 (Frichot and François 2015). This approach uses the results of ancestry estimation from a sparse non-negative matrix
265 factorization (sNMF) analysis to infer missing genotypes (Frichot et al. 2014). In sNMF, we set K to infer four clusters
266 and kept the best out of 10 runs based on a cross entropy criterion.

267 Phenotyping of flowering time and panicle morphology

268 Phenotyping of flowering time and panicle morphology was performed near Banfora (Burkina-Faso) under irrigated
269 field conditions at the Institut de l’Environnement et de Recherches Agricoles (INERA) station in 2012 and 2014.
270 Plants were sown at two different periods in the same year: the first at beginning of June (“early sowing”) and second in
271 mid-July (“late sowing”). A total of 15 plants per plot of 0.5 m² were grown. The field trials followed an alpha-lattice
272 design with two replicates (Patterson and Williams 1976) per date of sowing per year. Each single block included 19
273 accessions (i.e. 19 plots). In total, 87 *O. glaberrima* accessions were planted in 2012 and 155 in 2014.

274 Flowering date (DFT) was scored when 50% of the plants for a given accession harbored heading panicles. Fourteen
275 days after heading date, the three main panicles from three central plants per plot per repeat were collected (i.e. nine
276 panicles/accession/repeat) from the early sowing. Each panicle was fixed on a white paper board, photographed and
277 phenotyped using the P-TRAP software (AL-Tam et al. 2013). We estimated rachis length (RL), primary branch
278 number (PBN), primary branch average length (PBL), primary branch internode average length (PBintL), secondary

279 branch number (SBN), secondary branch internode average length (SBintL) and spikelet number (SpN). All statistical
280 analyses of the dataset were performed using R packages ade4, corrplot and agricolae (R software version 1.2.1335) as
281 described in Ta et al. (2018).

282 **RYMV resistance phenotyping**

283 Resistance was evaluated based on ELISA performed on infected plants cultivated in the greenhouse, under controlled
284 conditions. As high resistance to RYMV is widely accepted for African rice, we excluded highly resistant accessions,
285 i.e. in which no virus can be detected with ELISA (Thiémmélé et al. 2010; Orjuela et al. 2013; Pidon et al. 2017), and we
286 focused only on quantitative resistance. We therefore assessed resistance on a set of 125 accessions. Two varieties were
287 used as susceptibility controls, IR64 (*O. sativa ssp. indica*) and Nipponbare (*O. sativa ssp. japonica*), and one as a high
288 resistance control, Tog5681 (*O. glaberrima*). Three replicate experiments of all varieties were performed. In each
289 experiment, plants were organized in two complete blocks with four plant replicates per accession.

290 Plants were mechanically inoculated three weeks after sowing, as described in Pinel-Galzi et al. (2018) with CI4 isolate
291 of RYMV (Pinel et al. 2000). Four discs of 4 mm diameter were cut on the last emerged leaf of each plant 17 and 20
292 days after inoculation (dai) and discs from the four plants of the same repeat were pooled. Samples were ground with a
293 QIAGEN TissueLyser II bead mill and resuspended in 750 μ L 1X PBST (Phosphate buffer saline with Tween 20).
294 Virus content was estimated by DAS-ELISA (Pinel-Galzi et al. 2018). Preliminary tests on a subset of samples were
295 performed to assess the dilution that best discriminated between samples. ELISA tests were finally performed at
296 dilutions of 1/1,000 for 17 dai sampling date and 1/2,500 for 20 dai sampling date. As virus content was highly
297 correlated between 17 and 20 days after infection ($R^2=0,81$), the resistance level was estimated as the mean of the two
298 sampling dates.

299 **Environmental variables**

300 For accessions with geographical sampling coordinates, we retrieved information for 19 climate-related variables
301 (referred to here as bio) from the WORLDCLIM database at a 2.5 minute resolution (Hijmans et al. 2005). We also
302 retrieved the average monthly maximum temperature (referred to here as Tmax). We first performed a Principal
303 Component Analysis (PCA) on each set of variables to build uncorrelated composite variables. PCA were performed
304 using R software (Frichot and François 2015). Association studies were performed using the first two components of
305 each PCA.

306 **Association studies**

307 For each trial, SNPs displaying a minimal allele frequency (frequency of the minor allele) lower than 5% were filtered
308 out. We first adjusted a simple linear model (Analysis of variance, ANOVA) to associate phenotype and genotype. This
309 simple method did not take into account any putative confounding factor and allowed us to assess whether taking into
310 account relatedness and/or population structure could reduce false positive rates. Two classes of methods accounting for
311 confounding factors were used: 1) mixed models using kinship matrix and/or population structure (Yu et al. 2006); and
312 2) latent factor methods (Frichot et al. 2013). We used both mixed linear models MLM (Zhang et al. 2010) as
313 implemented in GAPIT R package (Lipka et al. 2012) and EMMA (Kang et al. 2008) as implemented in R package
314 EMMA. For EMMA, the kinship matrix was estimated using the emma.kinship function. For MLM (Q+K model), the
315 kinship (K matrix) was computed using the Van Raden method and the first three principal components of a PCA of
316 genomic data were used as the Q matrix. Finally, we used latent factor methods (Frichot et al. 2013) that jointly
317 estimated associations between genotype and phenotype and confounding factors. We used the R packages LFMM2
318 (Caye et al. 2019) and CATE (Wang et al. 2017) to perform these analyses. For LFMM2, we first made the estimation
319 of the confounding factors by using a subset of SNPs obtained by applying a 20% MAF filter, and we considered four
320 latent factors (Cubry et al. 2018). We then used the resulting confounding matrix for the analysis of genotype/phenotype
321 association. For CATE, we considered all SNPs and we assumed four confounding factors in the association model. The
322 results of all analyses were graphically represented by using a QQ-plot to assess confounding factor correction and
323 Manhattan plots (R package qqman, Turner 2014). We used a false discovery rate (FDR) of 5% to select candidate
324 SNPs for each method. FDR estimation was realized using the R package qvalue (Storey et al. 2019).

325 GWAS analysis was performed separately for each year and trial (see Additional file 1: Table S1). *P*-values obtained
326 for the same traits or the same planting data were combined across experiments using Fisher's method (Sokal and Rohlf
327 2012). The final list of candidate SNPs was established for each trait by considering SNPs detected by at least one
328 method. Annotation of retained candidate SNPs was performed using the SNPeff annotation data for MSU7 (Kawahara
329 et al. 2013), considering genes within the region 25kb upstream and 25kb downstream from each detected SNP. We
330 used the funRiceGenes database (<http://funricegenes.ncpgr.cn>; Yao et al. 2018) to extract gene symbols and associated
331 functional information whenever possible.

332 Finally, for flowering traits, we established a list of known genes of particular interest from published data (Tsuji et al.
333 2011; Hori et al. 2016). This "expert" list was then used to assess the performance of our GWAS approach to retrieve
334 these potential candidates. We used a G-test to assess enrichment of candidates in our list of identified genes.

335 Gene ontology (GO) term enrichment tests for biological process, cellular component and molecular function terms
336 were performed using the Fisher exact test implemented in the R package TopGO (Alexa et al. 2006).

337

338 **List of abbreviations**

339 ANOVA – Analysis of Variance

340 BioPC1 – First principal component of PCA on bioclimatic variables

341 BioPC2 – Second principal component of PCA on bioclimatic variables

342 dai – Days after inoculation

343 DFT – Days to flowering

344 ELISA – Enzyme-linked immunosorbent assay

345 FDR – False discovery rate

346 GO – Gene ontology

347 GWAS – Genome wide association study

348 INERA – Institut de l'environnement et de recherches agricoles

349 LD – Long days

350 NLR – Nucleotide binding leucine rich repeat

351 PbintL – Primary branch internode average length

352 PBL – Primary branch average length

353 PBN – Primary branch number

354 PCA – Principle component analysis

355 QTL – Quantitative trait locus

356 QQ-plot – quantile-quantile plot

357 RL – Rachis length

358 RYMV – Rice yellow mottle virus

359 SbintL – Secondary branch internode average length

360 SBN – Secondary branch number

361 SD -Short days

362 SpN – Spikelet number

363 SNP – Single nucleotide polymorphism

364 TF – Transcription factor

365 TmaxPC1 – First principal component of PCA on maximal temperature variables

366 TmaxPC2 – Second principal component of PCA on maximal temperature variables

367 **Declarations**

368 • **Ethics approval and consent to participate**

369 Not applicable

370 • **Consent for publication**

371 Not applicable

372 • **Availability of data and material**

373 All customized R scripts are available as a GitHub repository: https://github.com/Africrop/gwas_african_rice. The
374 repository also contains the imputed genotypic data used here to reproduce exactly the same analysis. Phenotypic data
375 are provided as supplemental material.

376 • **Competing interests**

377 The authors declare that they have no competing interests.

378 • **Funding**

379 This work was supported by a grant from the France Génomique French National infrastructure and funded as part of
380 “Investissement d’avenir” (ANR-10-INBS-09) and the IRIGIN project (<http://irigin.org>) to FS, an ANR grant (ANR-
381 13-BSV7-0017) to YV, and a grant from Agropolis Foundation (through the « Investissements d’avenir » programme
382 (ANR-10-LABX-0001- 01) and Fondazione Cariplo under the reference ID EVOREPRICE 1201–004 to SJ. YV is also
383 supported by the Agropolis Resources Center for Crop Conservation, Adaptation and Diversity (ARCAD) with support
384 from the European Union FEDER program and from the Agropolis Foundation. PC was supported by an ANR grant
385 (AfriCrop project, ANR-13-BSV7-0017). The French Ministère de l’Enseignement Supérieur et de la Recherche
386 provided a PhD grant to HP. MH was supported by the CGIAR research program on rice.

387 • **Authors' contributions**

388 YV, LA, SJ and FS planned and supervised the study. HP, TKN, SJ, LA, HA, HK, HC and AG participated in data
389 generation and collection management. PC, TKN, HP, CTD, ACT, MH, OF, FS, YV, SJ and LA participated in the
390 statistical analysis. PC, HP, FS, YV, LA and SJ wrote the manuscript. All the authors read and agreed the manuscript.

391 • **Acknowledgements**

392 We thank Ndomassi Tando and the IRD itrop “Plantes Santé” bioinformatic platform for providing HPC resources and
393 support for our research project. We thank Mikael Valéro for technical help in RYMV resistance evaluation. The
394 authors thank staff from INERA station at Banfora (Burkina Faso) for support for field experiment and phenotype
395 scoring, and Ha Thi Loan from LMI RICE (Vietnam) for panicle trait scoring. We also thank Dr James Tregear, IRD,
396 for his careful reading and language editing of the manuscript.

397 REFERENCES

- 398 Agrawal GK, Abe K, Yamazaki M, et al (2005) Conservation of the E-function for Floral Organ Identity in Rice
399 Revealed by the Analysis of Tissue Culture-induced Loss-of-Function Mutants of the OsMADS1 Gene. *Plant*
400 *Mol Biol* 59:125–135. <https://doi.org/10.1007/s11103-005-2161-y>
- 401 Albar L, Bangratz-Reyser M, Hébrard E, et al (2006) Mutations in the eIF(iso)4G translation initiation factor confer
402 high resistance of rice to *Rice yellow mottle virus*. *Plant J* 47:417–426. [https://doi.org/10.1111/j.1365-](https://doi.org/10.1111/j.1365-313X.2006.02792.x)
403 [313X.2006.02792.x](https://doi.org/10.1111/j.1365-313X.2006.02792.x)
- 404 Alexa A, Rahnenführer J, Lengauer T (2006) Improved scoring of functional groups from gene expression data by
405 decorrelating GO graph structure. *Bioinformatics* 22:1600–1607. <https://doi.org/10.1093/bioinformatics/btl140>
- 406 AL-Tam F, Adam H, Anjos A dos, et al (2013) P-TRAP: a Panicle Trait Phenotyping tool. *BMC Plant Biol* 13:122.
407 <https://doi.org/10.1186/1471-2229-13-122>
- 408 Bai X, Zhao H, Huang Y, et al (2016) Genome-wide association analysis reveals different genetic control in panicle
409 architecture between *indica* and *japonica* Rice. *Plant Genome* 9.
410 <https://doi.org/10.3835/plantgenome2015.11.0115>
- 411 Boissard A, Albar L, Thiéméle D, et al (2007) Evaluation of genes from eIF4E and eIF4G multigenic families as
412 potential candidates for partial resistance QTLs to *Rice yellow mottle virus* in rice. *Theor Appl Genet* 116:53–
413 62. <https://doi.org/10.1007/s00122-007-0646-6>
- 414 Browning SR (2008) Missing data imputation and haplotype phase inference for genome-wide association studies. *Hum*
415 *Genet* 124:439–450. <https://doi.org/10.1007/s00439-008-0568-7>
- 416 Caye K, Jumentier B, Lepeule J, François O (2019) LFMM 2: Fast and Accurate Inference of Gene-Environment
417 Associations in Genome-Wide Studies. *Mol Biol Evol* 36:852–860. <https://doi.org/10.1093/molbev/msz008>
- 418 Choi JY, Zaidem M, Gutaker R, et al (2019) The complex geography of domestication of the African rice *Oryza*
419 *glaberrima*. *PLOS Genet* 15:e1007414. <https://doi.org/10.1371/journal.pgen.1007414>
- 420 Cosson P, Sofer L, Le QH, et al (2010) RTM3, which controls long-distance movement of potyviruses, is a member of a
421 new plant gene family encoding a meprin and TRAF homology domain-containing protein. *Plant Physiol*
422 154:222–232. <https://doi.org/10.1104/pp.110.155754>
- 423 Crowell S, Korniliev P, Falcão A, et al (2016) Genome-wide association and high-resolution phenotyping link *Oryza*
424 *sativa* panicle traits to numerous trait-specific QTL clusters. *Nat Commun* 7:1–14.
425 <https://doi.org/10.1038/ncomms10527>
- 426 Cubry P, Tranchant-Dubreuil C, Thuillet A-C, et al (2018) The rise and fall of African rice cultivation revealed by
427 analysis of 246 new genomes. *Curr Biol* 28:2274–2282.e6. <https://doi.org/10.1016/j.cub.2018.05.066>
- 428 Dametto A, Sperotto RA, Adamski JM, et al (2015) Cold tolerance in rice germinating seeds revealed by deep RNAseq
429 analysis of contrasting indica genotypes. *Plant Sci* 238:1–12. <https://doi.org/10.1016/j.plantsci.2015.05.009>
- 430 de Ronde D, Butterbach P, Kormelink R (2014) Dominant resistance against plant viruses. *Front Plant Sci* 5.
431 <https://doi.org/10.3389/fpls.2014.00307>

- 432 Frichot E, François O (2015) LEA: An R package for landscape and ecological association studies. *Methods Ecol Evol*
433 6:925–929. <https://doi.org/10.1111/2041-210X.12382>
- 434 Frichot E, Mathieu F, Trouillon T, et al (2014) Fast and efficient estimation of individual ancestry coefficients. *Genetics*
435 196:973–983. <https://doi.org/10.1534/genetics.113.160572>
- 436 Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for Associations between Loci and Environmental
437 Gradients Using Latent Factor Mixed Models. *Mol Biol Evol* 30:1687–1699.
438 <https://doi.org/10.1093/molbev/mst063>
- 439 Frichot E, Schoville SD, de Villemereuil P, et al (2015) Detecting adaptive evolution based on association with
440 ecological gradients: Orientation matters! *Heredity* 115:22–28. <https://doi.org/10.1038/hdy.2015.7>
- 441 Hayama R, Yokoi S, Tamaki S, et al (2003) Adaptation of photoperiodic control pathways produces short-day
442 flowering in rice. *Nature* 422:719–722. <https://doi.org/10.1038/nature01549>
- 443 Hayashi N, Inoue H, Kato T, et al (2010) Durable panicle blast-resistance gene *Pbl* encodes an atypical CC-NBS-LRR
444 protein and was generated by acquiring a promoter through local genome duplication. *Plant J* 64:498–510.
445 <https://doi.org/10.1111/j.1365-313X.2010.04348.x>
- 446 Hébrard E, Poulicard N, Gérard C, et al (2010) Direct interaction between the *Rice yellow mottle virus* (RYMV) VPg
447 and the central domain of the rice eIF(iso)4G1 factor correlates with rice susceptibility and RYMV virulence.
448 *Mol Plant Microbe Interact* 23:1506–1513. <https://doi.org/10.1094/MPMI-03-10-0073>
- 449 Hijmans RJ, Cameron SE, Parra JL, et al (2005) Very high resolution interpolated climate surfaces for global land
450 areas. *Int J Climatol* 25:1965–1978. <https://doi.org/10.1002/joc.1276>
- 451 Hori K, Matsubara K, Yano M (2016) Genetic control of flowering time in rice: integration of Mendelian genetics and
452 genomics. *Theor Appl Genet* 129:2241–2252. <https://doi.org/10.1007/s00122-016-2773-4>
- 453 Huang X, Kurata N, Wei X, et al (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature*
454 490:497–501. <https://doi.org/10.1038/nature11532>
- 455 Ikeda K, Ito M, Nagasawa N, et al (2007) Rice ABERRANT PANICLE ORGANIZATION 1, encoding an F-box
456 protein, regulates meristem fate. *Plant J* 51:1030–1040. <https://doi.org/10.1111/j.1365-313X.2007.03200.x>
- 457 Ikeda-Kawakatsu K, Yasuno N, Oikawa T, et al (2009) Expression Level of ABERRANT PANICLE
458 ORGANIZATION1 determines rice inflorescence form through control of cell proliferation in the meristem.
459 *Plant Physiol* 150:736–747. <https://doi.org/10.1104/pp.109.136739>
- 460 Issaka S, Basso A, Sorho F, et al (2012) Diagnosis and importance of rice yellow mottle disease epidemics in Niger
461 republic. *J Appl Biosc* 50:3501-3511
- 462 Jeon J-S, Lee S, Jung K-H, et al (2000) Production of transgenic rice plants showing reduced heading date and plant
463 height by ectopic expression of rice MADS-box genes. *Mol Breed* 6:581–592.
464 <https://doi.org/10.1023/A:1011388620872>
- 465 Jeong HJ, Yang J, Cho L-H, An G (2016) OsVIL1 controls flowering time in rice by suppressing OsLF under short
466 days and by inducing Ghd7 under long days. *Plant Cell Rep* 35:905–920. <https://doi.org/10.1007/s00299-015-1931-5>
- 468 Kam H, Laing MD, Ouoba J, Ndjioudjop MN (2013) Rice traits preferred by farmers and their perceptions of rice
469 yellow mottle virus (RYMV) disease in Cascades Region of Burkina Faso. *Afr J Agricult* 8:2703-2712.
470 <https://doi.org/10.5897/AJAR12.1723>
- 471 Kang HM, Zaitlen NA, Wade CM, et al (2008) Efficient control of population structure in model organism association
472 mapping. *Genetics* 178:1709–1723. <https://doi.org/10.1534/genetics.107.080101>
- 473 Kawahara Y, Bastide M de la, Hamilton JP, et al (2013) Improvement of the *Oryza sativa* Nipponbare reference
474 genome using next generation sequence and optical map data. *Rice* 6:4. <https://doi.org/10.1186/1939-8433-6-4>

- 475 Khanday I, Yadav SR, Vijayraghavan U (2013) Rice LHS1/OsMADS1 controls floret meristem specification by
476 coordinated regulation of transcription factors and hormone signaling pathways. *Plant Physiol* 161:1970–1983.
477 <https://doi.org/10.1104/pp.112.212423>
- 478 Kim SL, Lee S, Kim HJ, et al (2007) OsMADS51 is a short-day flowering promoter that functions upstream of Ehd1,
479 OsMADS14, and Hd3a. *Plant Physiol* 145:1484–1494. <https://doi.org/10.1104/pp.107.103291>
- 480 Kim S-M, Suh J-P, Lee C-K, et al (2014) QTL mapping and development of candidate gene-derived DNA markers
481 associated with seedling cold tolerance in rice (*Oryza sativa* L.). *Mol Genet Genomics* 289:333–343.
482 <https://doi.org/10.1007/s00438-014-0813-9>
- 483 Komiya R, Ikegami A, Tamaki S, et al (2008) Hd3a and RFT1 are essential for flowering in rice. *Development*
484 135:767–774. <https://doi.org/10.1242/dev.008631>
- 485 Kouassi NK, N'Guessan P, Albar L, et al (2005) Distribution and characterization of *Rice yellow mottle virus*: A threat
486 to African farmers. *Plant Disease* 59:124-133 <https://doi.org/10.1094/PD-89-0124>
- 487 Lee Y-S, An G (2015) Regulation of flowering time in rice. *J Plant Biol* 58:353–360. <https://doi.org/10.1007/s12374-015-0425-x>
- 489 Li S, Qian Q, Fu Z, et al (2009) Short panicle1 encodes a putative PTR family transporter and determines rice panicle
490 size. *Plant J* 58:592–605. <https://doi.org/10.1111/j.1365-313X.2009.03799.x>
- 491 Lipka AE, Tian F, Wang Q, et al (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics*
492 28:2397–2399. <https://doi.org/10.1093/bioinformatics/bts444>
- 493 Liu T, Liu H, Zhang H, Xing Y (2013) Validation and characterization of Ghd7.1, a major Quantitative Trait Locus
494 with pleiotropic effects on spikelets per panicle, plant height, and heading date in rice (*Oryza sativa* L.). *J*
495 *Integr Plant Biol* 55:917–927. <https://doi.org/10.1111/jipb.12070>
- 496 Marandel G, Pascal T, Candresse T, Decroocq V (2009) Quantitative resistance to *Plum pox virus* in *Prunus davidiana*
497 P1908 linked to components of the eukaryotic translation initiation complex. *Plant Pathol* 58:425–435.
498 <https://doi.org/10.1111/j.1365-3059.2008.02012.x>
- 499 Marchini J, Howie B (2010) Genotype imputation for genome-wide association studies. *Nat Rev Genet* 11:499–511.
500 <https://doi.org/10.1038/nrg2796>
- 501 Meyer RS, Choi JY, Sanches M, et al (2016) Domestication history and geographical adaptation inferred from a SNP
502 map of African rice. *Nat Genet* 48:1083–1088. <https://doi.org/10.1038/ng.3633>
- 503 Molla KA, Debnath AB, Ganie SA, Mondal TK (2015) Identification and analysis of novel salt responsive candidate
504 gene based SSRs (cgSSRs) from rice (*Oryza sativa* L.). *BMC Plant Biol* 15:122.
505 <https://doi.org/10.1186/s12870-015-0498-1>
- 506 Nakagawa M, Shimamoto K, Kyojuka J (2002) Overexpression of RCN1 and RCN2, rice TERMINAL FLOWER
507 1/CENTRORADIALIS homologs, confers delay of phase transition and altered panicle morphology in rice.
508 *Plant J* 29:743–750. <https://doi.org/10.1046/j.1365-313X.2002.01255.x>
- 509 Nicaise V, German-Retana S, Sanjuán R, et al (2003) The eukaryotic translation initiation factor 4E controls lettuce
510 susceptibility to the *Potyvirus lettuce mosaic virus*. *Plant Physiol* 132:1272–1282.
511 <https://doi.org/10.1104/pp.102.017855>
- 512 Ogiso E, Takahashi Y, Sasaki T, et al (2010) The Role of Casein Kinase II in Flowering Time Regulation Has
513 Diversified during Evolution. *Plant Physiol* 152:808–820. <https://doi.org/10.1104/pp.109.148908>
- 514 Omidbakhshfard MA, Proost S, Fujikura U, Mueller-Roeber B (2015) Growth-Regulating Factors (GRFs): A small
515 transcription factor family with important functions in plant biology. *Mol Plant* 8:998–1010.
516 <https://doi.org/10.1016/j.molp.2015.01.013>

- 517 Ookawa T, Hobo T, Yano M, et al (2010) New approach for rice improvement using a pleiotropic QTL gene for
518 lodging resistance and yield. *Nat Commun* 1:1–11. <https://doi.org/10.1038/ncomms1132>
- 519 Orjuela J, Thiémélé D, Kolade O, et al (2013) A recessive resistance to *Rice yellow mottle virus* is associated with a rice
520 homolog of the CPR5 gene, a regulator of active defense mechanisms. *Mol Plant Microbe Interact* 26:1455–63.
521 <https://doi.org/10.1094/MPMI-05-13-0127-R>
- 522 Orjuela J, Sabot F, Chéron S, et al (2014) An extensive analysis of the African rice genetic diversity through a global
523 genotyping. *Theor Appl Genet* 127:2211–2223. <https://doi.org/10.1007/s00122-014-2374-z>
- 524 Patterson HD, Williams ER (1976) A new class of resolvable incomplete block designs. *Biometrika* 63:83–92.
525 <https://doi.org/10.1093/biomet/63.1.83>
- 526 Pidon H, Ghesquière A, Chéron S, et al (2017) Fine mapping of RYMV3: a new resistance gene to *Rice yellow mottle*
527 *virus* from *Oryza glaberrima*. *Theor Appl Genet* 130:807–818. <https://doi.org/10.1007/s00122-017-2853-0>
- 528 Pinel A, N’Guessan P, Bousalem M, Fargette D (2000) Molecular variability of geographically distinct isolates of *Rice*
529 *yellow mottle virus* in Africa. *Arch Virol* 145:1621–1638. <https://doi.org/10.1007/s007050070080>
- 530 Pinel-Galzi A, Hébrard E, Traoré O, et al (2018) Protocol for RYMV inoculation and resistance evaluation in rice
531 seedlings. *Bio-Protoc* 8:e2863. <https://doi.org/10.21769/BioProtoc.2863>
- 532 Pitzalis N, Heinlein M (2018) The roles of membranes and associated cytoskeleton in plant virus replication and cell-to-
533 cell movement. *J Exp Bot* 69:117–132. <https://doi.org/10.1093/jxb/erx334>
- 534 Rebolledo MC, Peña AL, Duitama J, et al (2016) Combining Image Analysis, Genome Wide Association Studies and
535 Different Field Trials to Reveal Stable Genetic Regions Related to Panicle Architecture and the Number of
536 Spikelets per Panicle in Rice. *Front Plant Sci* 7. <https://doi.org/10.3389/fpls.2016.01384>
- 537 Saikumar S, Kalmeshwer Gouda P, Saiharini A, et al (2014) Major QTL for enhancing rice grain yield under lowland
538 reproductive drought stress identified using an *O. sativa/O. glaberrima* introgression line. *Field Crops Res*
539 163:119–131. <https://doi.org/10.1016/j.fcr.2014.03.011>
- 540 Sakamoto T, Matsuoka M (2004) Generating high-yielding varieties by genetic manipulation of plant architecture. *Curr*
541 *Opin Biotechnol* 15:144–147. <https://doi.org/10.1016/j.copbio.2004.02.003>
- 542 Sanfaçon H (2015) Plant translation factors and virus resistance. *Viruses* 7:3392–3419.
543 <https://doi.org/10.3390/v7072778>
- 544 Sarla N, Swamy BPM (2005) *Oryza glaberrima*: A source for the improvement of *Oryza sativa*. *Curr Sci* 89:955–963
- 545 Sokal RR, Rohlf FJ (2012) *Biometry: the principles and practice of statistics in biological research*, [Extensively rev.]
546 4th ed. W.H. Freeman, New York
- 547 Storey JD, Bass AJ, Dabney A, et al (2019) qvalue: Q-value estimation for false discovery rate control. *Bioconductor*
548 version: Release (3.9)
- 549 Swaminathan K, Peterson K, Jack T (2008) The plant B3 superfamily. *Trends Plant Sci* 13:647–655.
550 <https://doi.org/10.1016/j.tplants.2008.09.006>
- 551 Ta KN, Khong NG, Ha TL, et al (2018) A genome-wide association study using a Vietnamese landrace panel of rice
552 (*Oryza sativa*) reveals new QTLs controlling panicle morphological traits. *BMC Plant Biol* 18:282.
553 <https://doi.org/10.1186/s12870-018-1504-1>
- 554 Takano M, Inagaki N, Xie X, et al (2005) Distinct and Cooperative Functions of Phytochromes A, B, and C in the
555 Control of Deetiolation and Flowering in Rice. *Plant Cell* 17:3311–3325.
556 <https://doi.org/10.1105/tpc.105.035899>
- 557 Teo ZWN, Song S, Wang Y-Q, et al (2014) New insights into the regulation of inflorescence architecture. *Trends Plant*
558 *Sci* 19:158–165. <https://doi.org/10.1016/j.tplants.2013.11.001>

- 559 Terao T, Nagata K, Morino K, Hirose T (2010) A gene controlling the number of primary rachis branches also controls
560 the vascular bundle formation and hence is responsible to increase the harvest index and grain yield in rice.
561 *Theor Appl Genet* 120:875–893. <https://doi.org/10.1007/s00122-009-1218-8>
- 562 Thiémélé D, Boissard A, Ndjiondjop M-N, et al (2010) Identification of a second major resistance gene to *Rice yellow*
563 *mottle virus*, *RYMV2*, in the African cultivated rice species, *O. glaberrima*. *Theor Appl Genet* 121:169–179.
564 <https://doi.org/10.1007/s00122-010-1300-2>
- 565 Tsuji H, Taoka K, Shimamoto K (2011) Regulation of flowering in rice: two florigen genes, a complex gene network,
566 and natural variation. *Curr Opin Plant Biol* 14:45–52. <https://doi.org/10.1016/j.pbi.2010.08.016>
- 567 Turner SD (2014) qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv 005165.
568 <https://doi.org/10.1101/005165>
- 569 Wang J, Zhao Q, Hastie T, Owen AB (2017) Confounder adjustment in multiple hypothesis testing. *Ann Stat* 45:1863–
570 1894. <https://doi.org/10.1214/16-AOS1511>
- 571 Wang M, Yu Y, Haberer G, et al (2014) The genome sequence of African rice (*Oryza glaberrima*) and evidence for
572 independent domestication. *Nat Genet* 46:982–988. <https://doi.org/10.1038/ng.3044>
- 573 Wang Y, Li J (2011) Branching in rice. *Curr Opin Plant Biol* 14:94–99. <https://doi.org/10.1016/j.pbi.2010.11.002>
- 574 Wang Z-X, Yano M, Yamanouchi U, et al (1999) The *Pib* gene for rice blast resistance belongs to the nucleotide
575 binding and leucine-rich repeat class of plant disease resistance genes. *Plant J* 19:55–64.
576 <https://doi.org/10.1046/j.1365-3113.1999.00498.x>
- 577 Wuriyangan H, Zhang B, Cao W-H, et al (2009) The Ethylene Receptor ETR2 Delays Floral Transition and Affects
578 Starch Accumulation in Rice. *Plant Cell* 21:1473–1494. <https://doi.org/10.1105/tpc.108.065391>
- 579 Xing Y, Zhang Q (2010) Genetic and molecular bases of rice yield. *Annu Rev Plant Biol* 61:421–442.
580 <https://doi.org/10.1146/annurev-arplant-042809-112209>
- 581 Xu Y, Zong W, Hou X, et al (2015) OsARID3, an AT-rich interaction domain-containing protein, is required for shoot
582 meristem development in rice. *Plant J* 83:806–817. <https://doi.org/10.1111/tj.12927>
- 583 Yao W, Li G, Yu Y, et al (2018) FunRiceGenes dataset for comprehensive understanding and application of rice
584 functional genes. *Gigascience* 7(1):1-9. <https://doi.org/10.1093/gigascience/gix119>
- 585 Yamaji Y, Maejima K, Komatsu K, et al (2012) Lectin-mediated resistance impairs plant virus infection at the cellular
586 level. *Plant Cell* 24:778–793. <https://doi.org/10.1105/tpc.111.093658>
- 587 Yano K, Yamamoto E, Aya K, et al (2016) Genome-wide association study using whole-genome sequencing rapidly
588 identifies new genes influencing agronomic traits in rice. *Nat Genet* 48:927–934.
589 <https://doi.org/10.1038/ng.3596>
- 590 Yano K, Morinaka Y, Wang F, et al (2019) GWAS with principal component analysis identifies a gene
591 comprehensively controlling rice architecture. *Proc Natl Acad Sci U S A*. 116:21262–21267.
592 <https://doi.org/10.1073/pnas.1904964116>
- 593 Yu J, Pressoir G, Briggs WH, et al (2006) A unified mixed-model method for association mapping that accounts for
594 multiple levels of relatedness. *Nat Genet* 38:203–208. <https://doi.org/10.1038/ng1702>
- 595 Zhang Z, Ersoz E, Lai C-Q, et al (2010) Mixed linear model approach adapted for genome-wide association studies. *Nat*
596 *Genet* 42:355–360. <https://doi.org/10.1038/ng.546>
- 597 Zhao K, Tung C-W, Eizenga GC, et al (2011) Genome-wide association mapping reveals a rich genetic architecture of
598 complex traits in *Oryza sativa*. *Nat Commun* 2:467. <https://doi.org/10.1038/ncomms1467>

599 **Table 1. Numbers of significant SNPs, regions and genes found to be associated with the different traits.**
 600 Significant SNPs were detected with the different models using the Fisher combination method and based on a FDR 5%
 601 threshold. The "All methods" column indicates the number of SNPs detected with at least one method. Fifty kb
 602 windows around these SNPs defined independent genomic regions associated with each trait.
 603

Trait	Significant SNPs					Regions	Genes
	EMMA	GAPIT	LFMM	CATE	All methods		
BioPC1	0	0	0	0	0	0	0
BioPC2	47	0	38	42	52	17	128
TmaxPC1	0	0	0	0	0	0	0
TmaxPC2	2,308	78	1,559	1,656	2,378	228	1,946
Early Sowing	850	0	1,343	1,044	1,450	80	732
Late Sowing	251	0	2,042	2,535	2,649	134	1,223
RYMV	2,171	387	197	156	2,264	257	2,248
SpN	0	0	0	0	0	0	0
PBN	0	0	0	257	257	25	243
SBN	1	0	0	0	1	1	10
RL	243	0	0	0	243	38	361
PBL	0	0	56	54	57	12	91
PBintL	1	0	0	1	1	1	9
SBintL	123	0	281	454	455	30	266

Table 2. List of the main candidate genes identified in regions associated with RYMV resistance, flowering time, panicle architecture and climate variation.

Trait	Region	Chrom.	Locus name (MSU)	Gene symbol	Annotation	Function	References
RYMV resistance	rOg-RYMV-77	11	LOC_Os11g43700		RGH1A_C_putative_C_expressed	candidate for resistance to RYMV, NLR gene	Pidon et al. 2016 doi: 10.1007/s00122-017-2853-0
	rOg-RYMV-159	4	LOC_Os04g42140	<i>Rymv1/eIF(iso)4G</i>	eukaryotic_initiation_factor_iso-4F_subunit_p82-34_C_putative_C_expressed	resistance to RYMV	Albar et al. 2006 doi: 10.1111/j.1365-313X.2006.02792.x
Flowering time	rOg-Flower-2	1	LOC_Os01g08700	<i>OsGI</i>	GIGANTEA_C_putative_C_expressed	Circadian rhythm/Flowering activator	Hayama et al. 2002 doi: 10.1093/pcp/pcf059
	rOg-Flower-29	1	LOC_Os01g69850	<i>OsMADS51</i>	OsMADS65_-_MADS-box_family_gene_with_MIKC_A_type-box_C_expressed	SD activator	Kim et al. 2007 doi: 10.1104/pp.107.103291
	rOg-Flower-71	3	LOC_Os03g54160	<i>OsMADS14</i>	OsMADS14_-_MADS-box_family_gene_with_MIKCc_type-box_C_expressed	Floral transition	Kim et al. 2007 doi: 10.1104/pp.107.103291
	rOg-Flower-71	3	LOC_Os03g54084	<i>OsPhyC</i>	phytochrome_C_C_putative_C_expressed	LD repressor	
	rOg-Flower-72	3	LOC_Os03g55389	<i>Hd6/CK2α</i>	casein_kinase_II_subunit_alpha-1_C_putative_C_expressed	LD repressor	Ogiso et al. 2010 doi: 10.1104/pp.109.148908
	rOg-Flower-83	4	LOC_Os04g08740	<i>ETR2</i>	ethylene_receptor_C_putative_C_expressed	Ethylene receptor	Wuriyangan et al. 2009 doi: 10.1105/tpc.108.065391
	rOg-Flower-114	6	LOC_Os06g06320	<i>Hd3a</i>	osFTL2_FT-Like2_homologous_to_Flowering_Locus_T_gene	Florigen	Kojima et al. 2002 doi: 10.1093/pcp/pcf156
	rOg-Flower-114	6	LOC_Os06g06300	<i>RFT1</i>	osFTL3_FT-Like3_homologous_to_Flowering_Locus_T_gene	Florigen	Kojima et al. 2002 doi: 10.1093/pcp/pcf156
rOg-Flower-168	11	LOC_Os11g05470	<i>RCN1</i>	RCN1_Centroradialis-like1_homogous_to_TFL1_gene	Similar to phosphatidylethanolamine-binding protein	Nakagawa et al. 2002 doi: 10.1046/j.1365-313X.2002.01255.x	
Panicle	rOg-PAN-22	3	LOC_Os03g11614	<i>OsMADS1/LH S1/AFO</i>	OsMADS1_-_MADS-box_family_gene_with_MIKCc_type-box_C_expressed	Spikelet meristem determinancy and floral organ	Jeon et al. 2000 doi: 10.1105/tpc.12.6.871

	rOg-PAN-59	6	LOC_Os06g45460	<i>APO1/OsAPO1/SCM2</i>	OsFBX202_-_F-box_domain_containing_protein_C_expressed	identity Inflorescence form, Loading resistance and grain yield	Ikeda et al. 2007	doi: 10.1111/j.1365-313X.2007.03200.x
	rOg-PAN-65	7	LOC_Os07g49460	<i>Ghd7.1/Hd2/OsPRR37/DT H7</i>	response_regulator_receiver_domain_containing_protein_C_expressed	Heading date, inflorescence form and grain yield	Liu et al. 2013	doi: 10.1111/jipb.12070
	rOg-PAN-91	11	LOC_Os11g12740	<i>SPI</i>	peptide_transporter_PTR2_C_putative_C_expressed	Panicle elongation	Li et al. 2009	doi: 10.1111/j.1365-313X.2009.03799.x.
Environment	rOg-BioClim-2	2	LOC_Os02g04630	<i>OsCAX4</i>	sodium_Fcalcium_exchanger_protein_C_putative_C_expressed	Salt tolerance gene	Molla et al. 2015	doi: 10.1186/s12870-015-0498-1
	rOg-BioClim-3	3	LOC_Os03g11614	<i>OsLG3b/OsMADS1/LHS1/AFO</i>	OsMADS1_-_MADS-box_family_gene_with_MIKCc_type-box_C_expressed	candidate for drought tolerance	Saikumar et al. 2014	doi: 10.1016/j.fcr.2014.03.011
	rOg-BioClim-4	3	LOC_Os03g53150	<i>OsIAA13/OsIAA1/OsAUX</i>	OsIAA13_-_Auxin-responsive_Aux_FIAA_gene_family_member_C_expressed	differentially expressed between cold-sensitive and cold-tolerant genotypes during seed germination	Dametto et al. 2015	doi: 10.1016/j.plantsci.2015.05.009

606 **FIGURE CAPTIONS**

607 **Figure 1. Manhattan plots of LFMM association results for flowering time assessed for early (upper pane) and**
608 **late (lower pane) sowing.** The red line indicates the 5% FDR threshold. Known Asian rice flowering genes found in
609 the vicinity of significant SNPs are indicated.

610
611 **Figure 2. Details of Manhattan plots obtained with LFMM on regions associated to RYMV resistance on**
612 **chromosome 4 (A) and 11 (B).** The 5% FDR threshold is represented by a red horizontal line. Positions of the major
613 resistance genes *RYMV1* and *RYMV3* are indicated by plain blue lines and positions of other *NLR* genes on chromosome
614 11 are indicated by dotted blue lines.

615 **ADDITIONAL FILE CAPTIONS**

616 **Additional file 1_Table S1. Phenotypic data used for genome-wide association analyses.** Flowering time (DFT),
617 rachis length (RL), primary branch number (PBN), primary branch average length (PBL), primary branch internode
618 average length (PBintL), secondary branch number (SBN), secondary branch internode average length (SBintL) and
619 spikelet number (SpN) were evaluated in field conditions in 2012 and 2014. Resistance to RYMV was evaluated in
620 greenhouse conditions during three experiments (RYMV1, RYMV2, RYMV3) shifted of about 1-2 months.
621 Environmental data were extracted from the worldclim database at available sampling locations. A Principal
622 Component Analysis was then performed on i) the whole set of variables and ii) only maximal temperature related ones.
623 The two first axes of both these PCA were then used for association analyses and are reported in this table.

624

625 **Additional file 2. Manhattan plots and QQ plots for flowering time.** Association analysis was performed on both
626 early and late plant sowings. The 5% FDR thresholds are indicated by red lines.

627

628 **Additional file 3. Manhattan plots and QQ plots for panicle architecture related traits.** Association analysis was
629 performed independently for each trait and repetition. *P*-values obtained for each replicate were then combined using a
630 Fisher method to obtain final *p*-values for each trait. The 5% FDR thresholds are indicated by red lines.

631

632 **Additional file 4. Manhattan plots and QQ plots for RYMV resistance.** The three replicates of phenotypic
633 evaluation were combined based on a Fisher method. The 5% FDR thresholds are indicated by red lines.

634

635 **Additional file 5. Manhattan plots and QQ plots for environmental data.** The 5% FDR thresholds are indicated by
636 red lines.

637

638 **Additional file 6_Table S2. List of the SNPs associated with eleven different phenotypic traits.** Significant SNPs
639 were identified based on four different models (EMMA, CATE, LFMM and GAPIT), the Fisher method to combine
640 several repetitions of phenotypic data and a 5% FDR threshold. For each trait, the *p*-values obtained with the different
641 models are indicated if significant.

642

643 **Additional file 7_Table S3. List of the genomic regions associated with five different categories of phenotypic**
644 **traits.** Regions were defined based on 50 kb windows around the significant SNPs detected with any of the four
645 models. Overlapping regions were combined into a single one. The chromosome (Chr), the starting (Position 1) and

646 ending (Position 2) positions, the size of the region (Intervals) in base pairs and the number of significant SNPs
647 included (Sign_SNPs_nb) are indicated. Sheets "RYMV", "Tmax", BioClim" concerned the regions identified for the
648 resistance to RYMV, the maximum temperature related variables and the whole set of bioclimatic variables,
649 respectively. The sheet "Flowering" concerned regions identified with the early sowing flowering time (Early) or the
650 late sowing flowering time (Late) traits, as indicated in the two last columns. The sheet "Panicle" concerned regions
651 identified with rachis length (RL), primary branch number (PBN), primary branch average length (PBL), primary
652 branch internode average length (PBintL), secondary branch number (SBN), secondary branch internode average length
653 (SBintL) and spikelet number (SpN), as indicated in the last columns.

654

655 **Additional file 8_Table S4. List of genes located in each of the regions associated with five categories of**
656 **phenotypic traits.**

657

658 **Additional file 9_Table S5. Expert list of genes previously described as involved in flowering time in Asian rice**
659 **and test of enrichment of the list of genes detected by our association analysis.** Fold-enrichment and G-test
660 associated *p*-value are reported.

661

662 **Additional file 10_Table S6. GO terms tables.** GO term analyses were run independently for each trait (TmaxPC2,
663 BioPC2, Flowering Time, Panicle, and RYMV). Results are shown for each trait and for the three GO term categories
664 (Biological processes, Cellular components, Molecular functions).



